View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 11, 151-159 (1965)

Partial Differential Equations for the Conditional Distribution of a Markov Process Given Noisy Observations

CHARLOTTE T. STRIEBEL

Lockheed Missiles and Space Division, Palo Alto, California

Submitted by Richard Bellman

1. INTRODUCTION

Let $x(t)$ be a vector valued continuous time parameter Markov process with transition probabilities

$$P_{t,\Delta}[x(t+\Delta)\in B \mid x(t)] \tag{1.1}$$

for all Borel sets B in the Euclidean n-space χ. Suppose that the process $x(t)$ cannot be observed directly but instead the process $\xi(t)$ is observed

$$\xi(t) = \int_0^t H(\tau)\, x(\tau)\, d\tau + \int_0^t K(\tau)\, dw(\tau) \tag{1.2}$$

where $w(t)$ is a separable Wiener process, more precisely a vector valued process whose components are independent and each is a separable Wiener process. The matrix valued functions $H(\tau)$ and $K(\tau)$ are known and defined in such a way that

$$R(\tau) = K(\tau)\, K(\tau)^* \tag{1.3}$$

is nonsingular for $0 \leqslant \tau \leqslant T$. If the process $x(t)$ is to be estimated instantaneously from accumulated observations, the estimation problem will be called the filtering problem, and the distribution of interest is

$$G(t, B) = P[x(t)\in B \mid \xi(\tau), \tau \leqslant t] \tag{1.4}$$

the conditional distribution of $x(t)$ given the observations up to time t. This is the distribution needed, for example, in the control problem in which action must be taken instantaneously, or in the case that the system is to be estimated only at the final time T. However, if the entire process $x(t)$, $0 \leqslant t \leqslant T$ is to be reconstructed after all observations have been taken or if it is to be estimated at a particular intervening time t', $0 \leqslant t' \leqslant T$. the estimation problem is termed a smoothing problem and the distribution of interest is

$$Q(t, B) = P[x(t)\in B \mid \xi(\tau), 0 \leqslant \tau \leqslant T]. \tag{1.5}$$

Partial differential equations will be given for both $G(t, B)$ and $Q(t, B)$. Rough derivations will be given indicating the line of reasoning to be followed. They will not be rigorous, and many of the details will be questionable. The assumptions required for the validity of the equations will be indicated, but will not be stated precisely.

2. The Discrete Case

The partial differential equations in the two cases will be derived from equations for the densities of the distributions in the discrete case by taking the limit as the time increment Δ goes to zero.

Let x_t, $t = 0, 1, 2, \cdots, T$ be a Markov process and let the observations be given by

$$\tilde{z}_t = z_t(x_t, \epsilon_t) \qquad t = 0, \cdots, T \tag{2.1}$$

where $z_t(\cdot, \cdot)$ is a known function and the ϵ_t are independent random variables, independent of the x_t process. Let Z_t denote the observations available at time t.

$$Z_t = \tilde{z}_0, \cdots, \tilde{z}_t. \tag{2.2}$$

The conditional density of x_{t+1} given Z_t is given by

$$\begin{aligned} p(x_{t+1} \mid Z_t) &= \int p(x_{t+1} \mid x_t, Z_t)\, g(x_t \mid Z_t)\, dx_t \\ &= \int p(x_{t+1} \mid x_t)\, g(x_t \mid Z_t)\, dx_t \end{aligned} \tag{2.3}$$

since x_{t+1} is independent of Z_t conditioned in x_t. This follows from the Markov property since Z_t depends on x_τ, $\tau = 0, \cdots, t$ and on the observation errors ϵ_τ, $\tau = 0, \cdots, t$ which are independent of x_t and x_{t+1}. The joint density of x_{t+1} and $\tilde{z}_{t+1}$ conditioned on Z_t is given by

$$\begin{aligned} p(x_{t+1}, \tilde{z}_{t+1} \mid Z_t) &= p(\tilde{z}_{t+1} \mid x_{t+1}, Z_t)\, p(x_{t+1} \mid Z_t) \\ &= p(\tilde{z}_{t+1} \mid x_{t+1})\, p(x_{t+1} \mid Z_t) \end{aligned} \tag{2.4}$$

since from (2.1) the distribution of $\tilde{z}_{t+1}$ is determined by x_{t+1}. The density of x_{t+1} conditioned on Z_t and $\tilde{z}_{t+1}$ can now be computed using (2.3) and (2.4)

$$\begin{aligned} g(x_{t+1} \mid Z_t, \tilde{z}_{t+1}) &= \frac{p(x_{t+1}, \tilde{z}_{t+1} \mid Z_t)}{\int p(x_{t+1}, \tilde{z}_{t+1} \mid Z_t)\, dx_{t+1}} \\ &= \frac{p(\tilde{z}_{t+1} \mid x_{t+1}) \int p(x_{t+1} \mid x_t)\, g(x_t \mid Z_t)\, dx_t}{\int p(\tilde{z}_{t+1} \mid x_{t+1})\, p(x_{t+1} \mid x_t)\, g(x_t \mid Z_t)\, dx_t dx_{t+1}}. \end{aligned} \tag{2.5}$$

The letter g is used instead of p for the densities of x_t given Z_t and x_{t+1} given Z_{t+1} since in the limit they will be the densities of the filtering distributions given by (1.4).

The density of x_t conditioned on Z_T is given by

$$\begin{aligned} q(x_t \mid Z_T) &= \int p(x_t \mid x_{t+1}, Z_t, \tilde{z}_{t+1}, \cdots, \tilde{z}_T)\, q(x_{t+1} \mid Z_T)\, dx_{t+1} \\ &= \int p(x_t \mid x_{t+1}, Z_t)\, q(x_{t+1} \mid Z_T)\, dx_{t+1} \end{aligned} \tag{2.6}$$

from the Markov property. Dividing numerator and denominator by the density of Z_t

$$p(x_t \mid x_{t+1}, Z_t) = \frac{p(x_t, x_{t+1}, Z_t)}{p(x_{t+1}, Z_t)} = \frac{p(x_t, x_{t+1} \mid Z_t)}{p(x_{t+1} \mid Z_t)}. \tag{2.7}$$

Using the Markov property again

$$\begin{aligned} p(x_t, x_{t+1} \mid Z_t) &= p(x_{t+1} \mid x_t, Z_t)\, g(x_t \mid Z_t) \\ &= p(x_{t+1} \mid x_t)\, g(x_t \mid Z_t). \end{aligned} \tag{2.8}$$

Substituting (2.8) into (2.7) and then (2.7) into (2.6)

$$q(x_t \mid Z_T) = g(x_t \mid Z_t) \int \frac{p(x_{t+1} \mid x_t)}{p(x_{t+1} \mid Z_t)}\, q(x_{t+1} \mid Z_T)\, dt_{t+1}. \tag{2.9}$$

The density $p(x_{t+1} \mid Z_t)$ can be obtained from $g(x_t \mid Z_t)$ by Eq. (2.3).

3. The Filtering Problem

In order to apply the formula (2.5) of the previous section let

$$\begin{aligned} x_t &= x(t\Delta) \\ \tilde{z}_t &= H(t)\, x(t\Delta) + K(t\Delta) \left[\frac{w(t\Delta) - w(t\Delta - \Delta)}{\Delta}\right]. \end{aligned} \tag{3.1}$$

Since $w(t)$ is a Wiener process the density of $\tilde{z}_{t+1}$ given x_{t+1} is normal with mean $H(t\Delta + \Delta)\, x_{t+1}$ and covariance $R(t\Delta + \Delta)/\Delta$. Let $f(x)$ be an arbitrary function of x and denote

$$(f, G_t^{\Delta}) = E[f(x_t) \mid \tilde{z}_0, \cdots, \tilde{z}_t] \tag{3.2}$$

Then from (2.5)

$$(f, G^{\Delta}_{(t+1)})$$

$$= \frac{\left(P_{t,\Delta}\left(f(x) \exp\left[-\frac{\Delta}{2}\left[\overset{*}{x}\overset{*}{H} R^{-1}Hx - \overset{*}{\tilde{z}}_{t+1}R^{-1}Hx - \overset{*}{x}\overset{*}{H} R^{-1}\tilde{z}_{t+1}\right]\right]\right), G_t^{\Delta}\right)}{\left(P_{t,\Delta}\left(\exp - \frac{\Delta}{2}[\ ]\right), G_t^{\Delta}\right)} \tag{3.3}$$

where $P_{t,\Delta}$ indicates the linear operator on the space of functions of x

$$P_{t,\Delta}(h(x))(y) = \int h(x)\, P_{t,\Delta}(dx \mid y)$$

$$= \int h(x_{t+1})\, p(x_{t+1} \mid y)\, dx_{t+1}. \tag{3.4}$$

The matrix functions H and R^{-1} in (3.3) are evaluated at $t\Delta + \Delta$. It will be assumed that for fixed t the family of linear operations $\{P_{t,\Delta}\}_{\Delta \geqslant 0}$ has an infinitesimal generator A_t, that is,

$$\frac{1}{\Delta}[(P_{t,\Delta}h, G) - (h, G)] \to (A_t h, G) \tag{3.5}$$

for the same class of functions $\mathscr{D}$ and a class of measures $\mathscr{G}$ which includes almost all $G(t, \cdot)$ given by (1.4). From (3.3)

$$\frac{1}{\Delta}[(f, G^{\Delta}_{(t+1)}) - (f, G_t^{\Delta})]$$

$$= \frac{\left(P_{t,\Delta} f \exp\left\{-\frac{\Delta}{2}[\ ]\right\}, G_t^{\Delta}\right) - (f, G_t^{\Delta})\left(P_{t,\Delta} \exp\left\{-\frac{\Delta}{2}[\ ]\right\}, G_t^{\Delta}\right)}{\Delta\left(P_{t,\Delta} \exp\left\{-\frac{\Delta}{2}[\ ]\right\}, G_t^{\Delta}\right)}. \tag{3.6}$$

Expanding the exponentials in the numerator to two terms and writing $P_{t,\Delta} = I + \Delta A_t$ where I is the identity operator

$$\frac{1}{\Delta}[(f, G^{\Delta}_{t+1}) - (f, G_t^{\Delta})]$$

$$= \frac{\left((I + \Delta A_t) f\left[1 - \frac{\Delta}{2}[\ ]\right], G_t^{\Delta}\right) - (f, G_t^{\Delta})\left((I + \Delta A_t)\left[1 - \frac{\Delta}{2}[\ ]\right], G_t^{\Delta}\right)}{\Delta\left(P_{t,\Delta} \exp\left[-\frac{\Delta}{2}[\ ]\right], G_t^{\Delta}\right)}$$

$$\to (A_t f, G(t)) - (\tfrac{1}{2} f[\ ], G(t)) - (f, G(t))\,(A_t 1, G(t)) + (f, G(t))\,(\tfrac{1}{2}[\ ], G(t)) \tag{3.7}$$

where

$$[\,] = x^*H^*(t)\,R^{-1}(t)\,x - \dot{\xi}(t)\,R^{-1}(t)\,H(t)\,x - x^*H^*(t)\,R^{-1}(t)\,\dot{\xi}(t)$$

is treated as a function of x and

$$\dot{\xi}(t) = \lim_{\Delta\to 0} \tilde{z}_{t,\Delta} \tag{3.8}$$

from (3.1) and (1.2).

$$(A_t 1, G(t)) = 0 \tag{3.9}$$

since $(P_{t,\Delta}1, G) = 1$ for G a probability measure.
Thus

$$\begin{aligned}\frac{\partial}{\partial t}(f, G(t)) = (A_t f, G(t)) &- \tfrac{1}{2}(x^*H(t)\,R^{-1}(t)\,H(t)\,xf, G(t)) \\ &+ \tfrac{1}{2}(f, G(t))\,(x^*H(t)\,R^{-1}(t)\,H(t)\,x, G(t)) \\ &+ \dot{\xi}(t)\,[(R^{-1}(t)\,H(t)\,xf, G(t)) \\ &- (f, G(t))\,(R^{-1}(t)\,H(t)\,x, G(t))]\end{aligned} \tag{3.10}$$

for $f \in \mathscr{D}$. This, of course, glosses over many difficulties.

The nonlinear terms of (3.10) have the effect of normalizing the measure to a probability measure. It can be verified directly that if a family of measures $F(t)$ satisfy the linear equation

$$\begin{aligned}\frac{\partial}{\partial t}(f, F(t)) = (A_t f, F(t)) &- \tfrac{1}{2}(x^*H(t)\,R^{-1}(t)\,H(t)\,xf, F(t)) \\ &+ \dot{\xi}^*(t)\,(R^{-1}(t)\,H(t)\,xf, F(t))\end{aligned} \tag{3.11}$$

then

$$G(t, B) = F(t, B)/(1, F(t)) \tag{3.12}$$

satisfies (3.10).

4. The Smoothing Distribution

The continuous case is related to the discrete case by Eqs. (3.1) in the previous section. From (2.9)

$$Q_t^\Delta = L_{t,\Delta}^* Q_{(t+1)}^\Delta \tag{4.1}$$

where $L_{t,\Delta}^*$ is the linear operator on the space of measures defined by

$$(L_{t,\Delta}^* Q)(B) = \int_{x_t \in B} \int_{X_{t+1}} \frac{p(x_t \mid \tilde{z}_0, \cdots, \tilde{z}_t)\, p(x_{t+1} \mid x_t)\, Q(dx_{t+1})\, dx_t}{p(x_{t+1} \mid \tilde{z}_0, \cdots, \tilde{z}_t)}. \tag{4.2}$$

The adjoint operator $L_{t,\Delta}$ satisfies

$$(L_{t,\Delta}f)(x_{t+1}) = \int f(x_t)\, p(x_t \mid \tilde{z}_0, \cdots, \tilde{z}_t) \frac{p(x_{t+1} \mid x_t)}{p(x_{t+1} \mid \tilde{z}_0, \cdots, \tilde{z}_t)}\, dx_t \tag{4.3}$$

for arbitrary f. Thus for arbitrary f and h

$$\int (L_{t,\Delta}f)(x_{t+1})\, h(x_{t+1})\, p(x_{t+1} \mid \tilde{z}_0, \cdots, \tilde{z}_t)\, dx_{t+1}$$
$$= \int \int f(x_t)\, p(x_t \mid \tilde{z}_0, \cdots, \tilde{z}_t)\, p(x_{t+1} \mid x_t)\, h(x_{t+1})\, dx_t dx_{t+1}\,. \tag{4.4}$$

This equation can be rewritten

$$(hL_{t,\Delta}f, P^*_{t,\Delta}G_t^{\Delta}) = (fP_{t,\Delta}h, G_t^{\Delta}) \tag{4.5}$$

where $P^*_{t,\Delta}$ is the linear operator

$$(P^*_{t,\Delta}G)(B) = \int_{x_{t+1}\epsilon B} \int_{X_t} p(x_{t+1} \mid x_t)\, G(dx_t) \tag{4.6}$$

which is adjoint to $P_{t,\Delta}$ given by (3.4).

It has already been assumed in (3.5) that

$$\frac{1}{\Delta}(P_{t,\Delta} - I) \to A_t\,. \tag{4.7}$$

If it is also assumed that

$$\frac{1}{\Delta}(L_{t,\Delta} - I) \to M_t\,. \tag{4.8}$$

Then from (4.5)

$$(hL_{t,\Delta}f, P^*_{t,\Delta}G_t^{\Delta}) = (P_{t,\Delta}(hL_{t,\Delta}f), G_t^{\Delta}) = ((I + \Delta A_t)(h(I + \Delta M_t) R), G_t^{\Delta})$$
$$(fP_{t,\Delta}h, G_t^{\Delta}) = (f(I + \Delta A_t)\, h, G_t^{\Delta}). \tag{4.9}$$

Hence

$$(A_t(hf), G(t)) + (hM_t f, G(t)) = (fA_t h, G(t)) \tag{4.10}$$

for f and h arbitrary. From (4.1)

$$\frac{1}{\Delta}[(f, Q_t^{\Delta}) - (f, Q_{t-1}^{\Delta})] = \frac{1}{\Delta}[(f, Q_t^{\Delta}) - (f, L^*_{t,\Delta}Q_t^{\Delta})]. \tag{4.11}$$

Thus

$$\frac{\partial}{\partial t}(f, Q(t)) = -(M_t f, Q(t)). \tag{4.12}$$

The right side of (4.12) can be computed from (4.10) by letting

$$h = q_t \tag{4.13}$$

the Radon Nikodyn derivative of $Q(t)$ with respect to $G(t)$. The result is

$$\frac{\partial}{\partial t}(f, Q(t)) = (A_t(fq_t), G(t)) - (PA_t q_t, G(t)). \tag{4.14}$$

5. The Gaussian Case

Consider the Markov process defined by

$$\dot{x} = \varphi x + \dot{\eta}(t) \tag{5.1}$$

where $\dot{\eta}(t)$ is white noise independent of $w(t)$ and with covariance

$$E(\dot{\eta}(s)\,\dot{\eta}(t)^*) = C(t)\,\delta(t - s) \qquad (5.2$$

and φ is a known matrix. For this process the operator A_t is given by

$$A_t f = \sum_i \sum_j \frac{\partial f(x)}{\partial x_i}\varphi_{ij}(t)\,x_j + \tfrac{1}{2}\sum_i \sum_j C_{ij}\frac{\partial^2 f(x)}{\partial x_i \partial x_j}. \tag{5.3}$$

Assuming that $G(t, \cdot)$ has a density $g(t, x)$ with respect to Lesbesque measure which has two derivatives, a partial differential equation for $g(t, x)$ is obtained from (3.10)

$$\begin{aligned}\frac{\partial}{\partial t} g(t, x) = &-\sum_i \sum_j \varphi_{ij} x_j \frac{\partial}{\partial x_i} g(t, x) - g(t, x)\sum_i \varphi_{ii} + \tfrac{1}{2}\sum_i \sum_j C_{ij}\frac{\partial^2 g(t, x)}{\partial x_i \partial x_j} \\ &- \tfrac{1}{2} g(t, x)\left[x^* H^* R^{-1} H x - \int x^* H^* R^{-1} H x g(t, x)\,dx\right] \\ &+ g(t, x)\,\dot{\xi}^*(t)\left[R^{-1} H x - \int R^{-1} H x g(t, x)\,dx\right].\end{aligned} \tag{5.4}$$

It can be verified by direct substitution that the solution to this equation is

$$g(\mathrm{t}, x) = \frac{1}{(2\pi)^{n/2}\,|\,P(t)\,|^{1/2}} \exp\left[-\tfrac{1}{2}(x - \hat{x}(t))^* P^{-1}(t)\,(x - \hat{x}(t))\right] \tag{5.5}$$

where $\hat{x}(t)$ and $P(t)$ satisfy the Kalman equations

$$\dot{\hat{x}} = \varphi\hat{x} + PH^*R^{-1}(\dot{\xi} - H\hat{x}) \tag{5.6}$$

$$\dot{P} = \varphi P + P\varphi^* - PH^*R^{-1}HP + C. \tag{5.7}$$

This solution (5.5) holds for an initial condition $g(0, x)$ Gaussian with mean $\hat{x}(0)$ and convariance $P(0)$. Equations (5.6) and (5.7) can be obtained directly from (3.10) if it assumed that $G(t, \cdot)$ is Gaussian and $\hat{x}$ and P are defined by

$$\hat{x}(t) = (x, G(t)) \tag{5.8}$$

$$P(t) = ((x - \hat{x}(t))\,(x - \hat{x}(t))^*, G(t)) \tag{5.9}$$

Let $q'(t, x)$ be the Lesbesque density of the distribution $Q(t, \cdot)$. Then

$$q(t, x) = \frac{q'(t, x)}{g(t, x)} \tag{5.10}$$

and a partial differential equation for q' can be obtained from (4.14) for A_t given by (5.3)

$$\frac{\partial}{\partial t} q'(t, x) = \frac{q'(t, x)}{g(t, x)} \Big[- \sum_i \sum_j \varphi_{ij} x_j \frac{\partial}{\partial x_i} g(t, x) - g(t, x) \sum_i \varphi_{ii} + \tfrac{1}{2} \sum_i \sum_j C_{ij} \frac{\partial^2}{\partial x_i \partial x_j} g(t, x)\Big]$$

$$- g(t, x) \Big[\sum_i \sum_j \varphi_{ij} x_j \frac{\partial}{\partial x_i} \Big(\frac{q'(t, x)}{g(t, x)}\Big) + \tfrac{1}{2} \sum_i \sum_j C_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \Big(\frac{q'(t, x)}{g(t, x)}\Big)\Big]. \tag{5.11}$$

For $g(t, x)$ given by (5.5), this equation is satisfied by

$$q'(t, x) = \frac{1}{(2\pi)^{n/2}\, |\, S(t)\, |^{1/2}} \exp\,[- \tfrac{1}{2}(x - \hat{\hat{x}}(t))^*\, S^{-1}(t)\, (x - \hat{\hat{x}}(t))] \tag{5.12}$$

where $\hat{\hat{x}}$ and S satisfy the equations for the smoothing problem

$$\dot{\hat{\hat{x}}} = \varphi\hat{\hat{x}} + CP^{-1}(\hat{\hat{x}} - \hat{x}) \tag{5.13}$$

$$\dot{S} = (\varphi + CP^{-1})\, S + S\, (\varphi + CP^{-1})^* - C \tag{5.14}$$

6. A Poisson Process

Let $x(t)$ be a Poisson process where the jumps are of unit size and the number of jumps in an interval of length t is a Poisson random variable with parameter λt. Let

$$f_n(y) = \begin{matrix} 1 & y = n \\ 0 & y \neq n. \end{matrix} \tag{6.1}$$

Then from the definition (3.5)

$$(A_t f_n)(x) = \lim \frac{1}{\Delta} \left[P([x(t+\Delta) = n \mid x(t) = x]) - f_n(x) \right]$$

$$= \begin{cases} \lambda & x = n-1 \\ -\lambda & x = n \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$

From (3.11) the set of equations

$$\dot{F}_n(t) = \lambda F_{n-1}(t) + \left[n\dot{\xi}(t) - \frac{n^2}{2} - \lambda \right] F_n(t) \qquad n = 0, 1, 2, \cdots$$

is obtained, where $F_n(t) = (f_n, F(t))$.

These equations can be solved iteratively in n, and the $G_n(t) = P[x(t) = n]$ are given by

$$G_n(t) = \frac{F_n(t)}{\sum_{n=0}^{\infty} F_n(t)}. \tag{6.3}$$

Again from the definition (3.5)

$$A(f_n q)(x) = \begin{cases} \lambda \dfrac{Q_n}{G_n} & x = n-1 \\ -\lambda \dfrac{Q_n}{G_n} & x = n \\ 0 & \text{otherwise} \end{cases}$$

and from (4.14)

$$\dot{Q}_n(t) = \lambda \left(Q_n \frac{G_{n-1}}{G_n} - Q_{n+1} \frac{G_n}{G_{n+1}} \right). \tag{6.5}$$

No convenient way is immediate for solving these equations.